TCAP/CRA Item Development, Review, and Scoring Process

What is a Constructed Response item?

- A Constructed Response (CR) item is any test item for which students must create a response.

- A CR item asks *more* than a multiple-choice item.

- A CR item often has more than one correct answer.

- A CR item assesses recall of information from long-term memory (multiple choice items only assess recognition), and allows students to show their thinking through their response.

What are the parts of a Constructed Response Assessment (CRA) item?

- Stem (the question or task)

- Sample response

- Rubric (the scoring guide)

You can see examples of these parts on the next three slides.

Sample Math Item Stem

Grade 7 SPI 0706.4.3 - Apply scale factor to solve problems involving area and volume.
CCSS 7.G.6 - Solve real-world and mathematical problems involving area, volume and surface area of two- and three-dimensional objects composed of triangles, quadrilaterals, polygons, cubes, and right prisms.

Marguerite has a section of wall that is 14" by 35". She is going to cover the wall with tiles that are each 2" by 5". How many of the tiles will be needed to cover the wall?

Explain how you determined your answer or show your work.

Sample Response

- The length of the tile is 1/7 the length of the wall, and the width of the tile is 1/7 the width of the wall. Since she has to cover the wall in two directions, she will need 7 tiles in each direction, which is 7 $\times$ 7 or 49 of the smaller tiles.

- I found the area of a tile, which is 10 in^2, and the area of the wall to be covered, which is 490 in^2 , and I divided 490 by 10, which is 49.

Rubric

- A rubric is an internal scoring guide for a CR item.

- A rubric assigns point values to clearly defined types of responses based on the amount of work required by the item. *The point value of an item is not related to the difficulty or the cognitive complexity (DOK level) of the item.*

- A rubric is only the *first* step in the development of a complete system for scoring large numbers of student responses fairly. Readers are trained to apply the rubric using students' papers. When readers score students' papers, they think more about the papers they trained on than the rubric they started with.

2-Point Math Rubric

2 points — The response indicates *complete* understanding of the task. The correct answer is given along with an explanation that demonstrates understanding.

1 point — The response indicates *partial* understanding of the task. The correct answer is given without an explanation of how it was found.

OR

The correct answer is not given, but the explanation demonstrates understanding.

0 points — The response is inaccurate.

OR

The response is unable to score.

Depth of Knowledge Levels

(Webb, 1997)

Each item on the CRA is assigned a DOK level based on the cognitive complexity of the item.

- Level 1 – Recall & Reproduction

requires recall of information; can involve following a simple procedure

- Level 2 – Skills & Concepts

requires application of information or knowledge; often more than one cognitive process/step

- Level 3 – Strategic Thinking

requires deep understanding exhibited through planning, using evidence, and more demanding cognitive reasoning

- Level 4 – Extended Thinking

requires high cognitive demand and is highly complex; most often requires an extended period of time

MI ◆ **MEASUREMENT**
INCORPORATED

What is the Importance of DOKs?
Depth of Knowledge Facts

- DOK levels can be used to ensure that the intent of the standard and the level of student demonstration required by that standard match the assessment item.

- The DOK level is not related to the point value or difficulty of an item.

- Verbs alone do not determine the cognitive complexity of assessment tasks.

- The DOK level is assigned based on the intent of an item.

How are CRA items developed?

- In the fall, TN teachers from across the state attend an item writing workshop and receive training in writing effective CRA items.

- At the same meeting a committee of R/LA teachers from across the state reviews passages and identifies those they feel are appropriate to include on the CRA; these passages will support the R/LA items.

- Over the next few months the teachers write items and submit them to Measurement Incorporated's Content Specialists, who provide feedback and work with the teachers to revise the items. This is to ensure that each item matches the standard being assessed, is worded clearly, and is unbiased (does not advantage or disadvantage any group of students).

How are CRA items reviewed? (part 1)

- Once an item has been accepted from the TN teacher who wrote it, it undergoes a rigorous internal review process.

- The item is reviewed by two Content Specialists.

- The item is reviewed by two editors for grammatical issues and to make sure the item adheres to the TCAP/CRA Style Guide.

- The item is reviewed by a scoring expert who looks for potential issues with the rubric that may cause problems during scoring.

- The item is reviewed by the Project Director.

- At any point in this process the item may return to the Content Specialist for revision, and the cycle repeats.

MI MEASUREMENT INCORPORATED

How are CRA items reviewed? (part 2)

- Once an item has been approved by Measurement Incorporated, it is presented to the State for review.
- The State reviews all items and additional revisions are made as needed.
- In the spring, TN teachers from across the state attend an item review workshop and receive training in reviewing CRA items.
- These teachers review all items that have been developed. They determine whether each item should be accepted, revised, or rejected.
- After the review workshop the Content Specialists make any requested revisions. Editors, scoring experts, and the Project Director review the revisions.
- The State reviews each item a final time before the items are considered approved. Approved items may now be field-tested.

MI MEASUREMENT INCORPORATED

How are CRA items scored?
Reader qualifications

- Every MI reader is required to have at least a bachelor's degree.
- A high percentage of MI readers have a graduate degree.
- Many MI readers are former or retired teachers or principals.
- Every reader must pass an initial screening prior to being selected for training (this includes a scored writing sample as well as a separate analytic component).

MI® **MEASUREMENT** INCORPORATED

How are CRA items scored?
Reader training

- After being selected for a project, readers are invited to train for the project.
- Training is comprehensive and provides readers with an opportunity to ask questions and build rapport with their team leader.
- For the CRA, readers must take and pass (80% correct) two qualifying sets of ten papers each to be hired to score for the project.
- Ongoing training is provided throughout the course of the project.

How are CRA items scored?
Scoring process

- Each student response is read by two different readers.
- If the two score points differ by more than one point, a third expert reader reads the student response. The third score becomes the official score.
- Responses that fall outside of the responses provided during training as part of the anchor papers are sent to the State for score decisions.
- Readers are monitored daily by MI staff. Every reader has a certain percentage of their responses checked daily.
- Both MI and State staff refer daily to scoring data to identify any teams or individual readers that need additional guidance or training.

MI ® **MEASUREMENT** INCORPORATED

Implications for Practice

- Use DOK levels in the classroom – encourage students to engage in deeper processing. The deeper students' processing, the better their understanding and memory. (Bruning, Schraw, & Norby, 2011)

- Incorporate CR items into assessments – classroom encoding and retrieval activities should match the contexts in which students ultimately will need to retrieve and use what they have learned. (Bruning, Schraw, & Norby, 2011)

- Adhere to best practices in assessment – this will allow for valid inferences to be drawn from your assessment data.